Voice portal hosting system and method

This application claims priority of European application EP00110430.6, filed on May 16, 2000, and whose contents are hereby incorporated by way of reference.

Field of the invention

The present invention concerns a voice portal hosting system and method. More specifically, the present invention concerns a portal system and method for hosting a plurality of voice applications and services from different external value-added service providers.

Related Art

The amount of information available in electronic format is large and growing at a fast rate worldwide. This information may be stored in computers and servers and accessed over communication networks such as e.g. the Internet. A popular portion of the Internet is the World Wide Web that allows for multimedia information such as e.g. text files, images, sound files, etc. to pass between servers and clients using the Hypertext Transfer Protocol (HTTP).

The clients may be desktop or palmtop computers or even mobile phones and use software programs, e.g. a browser, for displaying the requested information to the users. Even if computers and browsers are extremely widespread nowadays, there are many situations in which, for different kinds of reasons, their use is not appropriate or possible. Especially mobile users are often reluctant to carry a computer or even a palmtop on travel. Moreover, the use of browsers is too cumbersome for many people or for many uses, for example when only a very short piece of information such as, for example, the exact time, sport results or meteorological forecasts, is needed or for simple electronic transactions.

Users usually request information from an http-server using a request entered on a keyboard and interpreted by the browser. The use of keyboards in many situations, like while driving cars, is dangerous if not forbidden. Many requests contain an URL (Uniform Resource Locator) of the server, which for most users is hard to remember.

On the other hand, interactive voice response systems, often called IVR systems or simply voice servers, are known, in which audio information can be stored and accessed by a plurality of users with spoken requests transmitted through the telephone network and interpreted in the voice server. Voice servers have the advantage over http servers that the information can be accessed without any prior knowledge and without any special equipment from any telephone terminal in the world, using a normal phone number which can be found in conventional white or yellow phone books.

IVR systems usually comprise a speech recognition module, often called speech recognizer or simply recognizer, for recognizing and executing commands uttered by the user. Some IVR furthermore include a speaker identification or verification module for identifying or verifying the user's identity. Those systems allow the user to navigate through a so-called "voice menu" and to enter commands and requests which are to be executed by the server.

When compared to conventional human-operated teleservices dedicated to the delivery of goods or services, automated IVR systems allow cost-savings to be achieved by reducing the necessary human resources. Furthermore, those systems offer a better service to users, since they will only rarely hear a busy tone.

A fast and accurate speech recognizer is of unparalleled importance in order to expand the market space of voice servers and voice-based e-commerce applications. In conventional IVR implementations, each service provider operates its own IVR system. IVR systems that are accessed only infrequently by some users don't get a chance to adapt the speech recognition models of those occasional users, so that most speech recognizers used for voice-enabled e-commerce applications use speaker-independent speech recognition algorithms. On a typical word recognition task, speaker-independent speech recognition systems may have twice the error rate of speaker-dependant systems. Consequently, speech recognition performances in known voice servers usually remain poor.

Some IVR systems achieve better performance by using speaker-dependent speech models, which are usually learned during an enrollment session. However, most users would not accept having to spend the time necessary for a different enrollment session with each individual service provider.

Moreover, the direct recognition of the user's billing and delivering address was considered to be, for the time being, out of reach of the current speech recognition technology.

Brief Summary of the Invention

One aim of the invention is to provide a voice portal system and method with improved speech recognition performances over existing methods and systems.

One other aim is to provide a voice portal system and method which avoid the need for an explicit user enrollment in the case of simple services, and a system that offers an implicit adaptation technique that learns the user's habits, his speech accent, and, finally, the language model he uses.

One other aim is to provide a voice portal system and method which use speaker-dependant speech recognition and which may be used even for improving speech recognition performances with users who access a specific voice-enabled service only occasionally and who don't want to spend time for an enrollment session with this service.

One other aim is to provide a voice portal system and method that are easier and cheaper to set up, especially for small service providers, and which allow for good recognition performances even with rarely used services.

In accordance with one embodiment of the present invention, a voice portal hosting system with which a plurality of users in a first voice telecommunication network can establish a connection with a voice equipment comprising a memory in which a plurality of interactive voice response applications have been independently uploaded through a second telecommunication network by a plurality of independent value-added service providers. At least a plurality of said interactive voice response applications use a common speech recognition module run on said system.

The voice portal hosting system may be hosted by a telco operator that provides the intelligent IVR system to a plurality of independent value-added service providers.

One advantage is that user speech and/or language models can be stored in the hosting system and shared between the various service providers. Thus each service provider can benefit from the overall improvement of speech recognition and/or speaker identification performance.

By hosting a plurality of services in a central host, an attractive voice portal offering a comprehensive range of services and proposed goods can be set up.

In a preferred embodiment, the inventive voice portal delivers highly valued engagement services, for example directory assistance service 23, meteorological forecasts etc., in order to keep the user coming back for more and to create a relationship with this user.

Brief description of the drawing

Fig. 1 shows a diagram of a telecommunication system including one embodiment of the inventive voice portal hosting system.

Detailed Description

Fig. 1 shows a diagram of a telecommunication system including a voice portal hosting system 2. A plurality of users 1 can establish a voice connection with this voice portal 2 over a public telecommunication network 10, such as a public switched telecommunication network (PSTN), an integrated services data network (ISDN), a mobile telecommunication network, e.g. a GSM network, or a voice-enabled IP-Network (VoIP). The connection is established by selecting the voice portal 2 phone number on a user terminal equipment 11 such as a phone terminal. Preferably, the voice portal phone number is a free business number in the 0800 range or, in a variant embodiment, a charging number in the 0900 range.

The voice portal hosting system includes a dialogue manager (not shown) using a speech synthesizer (not shown) and a speech recognition module 24 (speech recognizer) for establishing a spoken dialogue with the user 1. The dialogue manager answers commands or requests uttered by the user with the requested information or service or by establishing a new connection with an external server. The speech recognition module preferably uses HMM (Hidden Markov Models), adaptive neural networks or hybrids networks. In a preferred embodiment, the speech recognition module comprises Hidden Markov Models built with a toolkit such as the HTK toolkit.

For each user registered in the system, a user profile is preferably stored in a database 22. The user profile may include user-specific information such as name, address, phone and fax numbers, e-mail address, preferred language, preferred delivering address, preferred billing address and/or type (e.g. credit card, check, bill, deposit account etc.), configuration preferences, ordering habits, a priori choices, interests and tastes and so on.

If the access to the portal system 2 is secured, the user profile may include security information such as e.g. passwords, biometric parameters, electronic public keys etc.

In a preferred embodiment, the voice portal hosting system 2 is operated by the operator of the telecommunication network 10. In this case, the user profile may include other information known from this operator such as for example solvability.

The system 2 further contains speech (and speaker) models and/or language models for each registered user; those models may be stored in databases 20, 21 or be implicitly contained in the speech recognizer 24.

The system 2 contains a user identification module 25 for determining the identity of the user 1 currently accessing the system. The module 25 may use the authentication CLI or IMSI (International Mobile Subscriber Identification) of the terminal 11 used by the user and/or a user identity entered on a keypad or uttered by the user for determining his identity. If the voice portal hosting system is operated by the operator of the telecommunication network 10, the A-caller identification may be used for that purpose even if this identification has been concealed for other purposes and for other subscribers, without violating the user's data protection rights. In a preferred embodiment, the user identification module 25 identifies in a first step the equipment 11 and uses in a second step a user identification method for determining which user, among the known users of the equipment 11, is currently calling the system. This feature allows for differencing among the several people in a household or in a company who share the same terminal equipment 11. In a variant embodiment, the user 1 is prompted to enter his user name when he calls the system; the declined identity is then checked using known speaker verification methods. The user identification module 25 preferably uses known HMM techniques and speaker models stored in the database 20.

Once the user has been identified by the module 25, his speech and language models as well as his profile can be retrieved from the databases 20, 21 and 22 and used by the dialogue manager (not shown) and by the speech recognizer 24 in order to adapt the dialogue and the speech recognition to each user. By using the fact that speaker-dependent speech recognition is used, a better performance is achieved and thus the quality of service will increase.

The one skilled in the art will understand that, instead of training and storing user-dependant speech models in the database 20, adaptation techniques can be applied. In this case, by using only a small amount of data from a new user, a good speaker-independent model set can be adapted using maximum likelihood linear regression to fit the characteristics of each user. This adaptation can proceed incrementally as adaptation data becomes available (incremental adaptation).

The voice portal hosting system 2 further preferably contains a directory assistance service 23 and/or various other engagement services which may be offered by the operator of the system 2 in order to attract a large number of users, to encourage them to use this system and to let them remember its phone number.

According to the invention, the voice portal hosting system 2 comprises a document server 26, which includes at least a memory, in which a plurality of voice applications can be independently uploaded through a telecommunication network 3 by a plurality of independent value-added service providers 4. Each application may e.g. correspond to an offer or a "voice site" of one provider; the user 1 can access each voice application from the common menu, which is offered to him when he calls the system.

At least several of the voice applications in the memory 26 use the common speech recognizer 24 run on the voice portal hosting system 2. In a preferred embodiment, several of the voice applications in the memory 26 share in the same way the language models 21 and/or at least part of the user profile 22.

The providers can upload a voice application in the system 3 through a telecommunication network, e.g. a TCP/IP network, a public switched telecommunication network (PSTN), an integrated services data network (ISDN), an X25 network, etc. The voice application may include a collection of software objects and/or components, such as executable files, Java-Beans, Corba components etc., which may be installed in the memory 26 and executed by a suitable processor. The voice application may comprise not-compiled or precompiled software modules, which are compiled or interpreted by a common compilation module (not shown) in said voice portal hosting system 2. In a preferred embodiment, the voice application comprises a VoiceXML document, or a set of VoiceXML documents, which form a conversational finite state machine and which are written using the known Voice extensible markup language (VoiceXML) established by the VoiceXML forum. The VoiceXML is a computer language designed to make information accessible via voice and phone and to create audio dialogs that feature synthesized speech, digitalized audio, recognition of spoken and DTMF inputs, recording of spoken inputs, and interactive exchanges. In this embodiment, the document server includes a VoiceXML interpreter for processing the VoiceXML documents requested by the users 1 calling the voice portal hosting system 2.

According to the invention, the recognizer 24 and/or the user identification module 25 make use of a common on-line speech adaptation module 240 for adapting the speech and/or language models, and/or for adapting the HMMs in the recognizer 24 and/or in the speaker identification module 25. This allows the large quantity of training material coming from all the underlying services 40, 41, 42, 43 etc. to be exploited The consequence is that the speech recognition is greatly enhanced by two possible procedures: (i) the personalization of the speech recognition to each user 1 or (ii) simply the increase of the quality of the speaker-independent speech recognition and speaker identification by periodic retraining with the on-field speech data. The goal of the first procedure is to adapt a speaker-independent speech recognition system to a particular user 1 and, in this way, to benefit from the increase of performance offered by speaker-dependent recognition methods. Additionally, the stochastic language model will be stored in a database 21 and updated for each specific user and service access. Therefore the antecedents of the language model can enhance the overall speech recognition result.

As a result, each service provider 40 can benefit for his VoiceXML documents from a recognizer using user speech models which are improved each time one user accesses any one of the different services 23, 40, 41, 42, 43 etc.

In speech recognition, the language model and the acoustic model are related by the Bayes' rule:

$$P(\omega_j | x) = p(x | \omega_j)\, P(\omega_j) \,/\, p(x)$$

where $P(\omega_j)$ is the language model, and $p(x | \omega_j)$ is the acoustic model: the probability density function of the observation sequence x given the class of vocabulary word ω_j. The classification is made based on the estimation of the maximum a posteriori probability $P(\omega_j | x)$.

Preferably, the voice portal hosting system 2 keeps listening to the communication while transferring the call to a specific value-added service provider 4; the purpose is to record the spoken items for the further adaptation of the speech and language models 20, 21.

In a preferred embodiment, users' calls are temporarily recorded in the voice portal hosting system 2, in order to provide for an off-line adaptation of the users' speech and language models. Calls may be erased after a predetermined time length or preferably as soon as they have been used for adapting the models.

Choices and selections made by the users in the voice menu are preferably individually and/or statistically stored, in order to provide for a general and/or user-specific improvement of the arborescence of the voice menu, or to provide for user-specific options and offers. Menu options, which are often selected by a specific user and/or by most users, may be reorganized in order to be more easily accessible. Moreover, the knowledge of the probability of each selection may be used for improving the quality of the speech recognition of each possible option in a menu.

The voice portal hosting system 2 preferably includes a common billing module and a common clearing center (not shown) for collecting and dispatching the collected amounts to the value-added service providers 4. As a result, a plurality of service providers can use shared modules, components, functions and/or objects in the voice portal hosting system 2 for billing the access to their services on a pay-per-minute or pay-per-view basis, and/or for billing the transactions between users 1 and providers 4. The billing can be effected with a common bill prepared by the operator of the voice portal hosting system 2 and grouping all the amounts due by the users to said operator and to the providers 4; if the system 2 is managed by the operator of the telecommunication network 10, the billing can be made with the phone bill sent to the network's users.

In a variant embodiment, at least several users 1 have a deposit account on said voice portal hosting system 2 which can be used for transactions with a plurality of value-added service providers. It is also possible to debit the users' credit card or phone card for each transaction or connection with any of the providers 4. Preferably, the billing address and/or preferences for each user are registered in the user profile database 22.

If the voice portal hosting system 2 integrates an access to the phone directory 23, as shown, the billing and/or delivering address of the user can also be found, checked or completed with this directory.

Using adapted requests, the various voice applications in the memory 26 can preferably access at least part of the content of the common user profile database 22 in order to retrieve e.g. the default billing address and preferences, the default delivery address, the user's preferred language and so on. In a preferred embodiment, at least part of the user data in the database 22 can be updated by a plurality of voice applications from different service providers.

At least some of the voice applications uploaded in the memory 26 are voice-enabled e-commerce applications that allow the voice portal hosting system 2 to "monetize" the relationship with the user 1. The example of a pizza automatic ordering system is a typical revenue-generating interactive voice response application. In that example, the service provider 4 is a pizza firm that pays a small amount per call or per order to the Telco operator and also a monthly fee for the service hosting.

In the following dialogue, the user 1 wishes to access the hosted voice application of the pizza company 40 for ordering a quattro stagioni pizza.

The residential user 1	**The voice portal hosting system 2 and the pizzeria 40:**
The user calls the voice portal hosting system 2 where the pizza e-shop is hosted.	The user identification module 25 has identified the user. The speech and language models 20, 21 are loaded.
	The voice portal hosting system: *Welcome, which e-shop would you like to visit?*
The user: *Give me the Blitz-Pizzas.*	The user's specific language 21 and speech 20 models are used to improve the recognition accuracy.
	hosted e-shop: *Welcome, what is your pizza choice?*
The user: *One Quattro stagioni, please!*	The "a priori" choice 22 of this user is known by the system and is given to the

	language model.
	hosted e-shop: *Is the following delivering and billing address correct? 44th Bakerstreet in London.*
The user: *That's correct!*	The following question is suggested by the user's profile 22.
	hosted e-shop: *Would you like an additional salad?*
The user: *No thanks!*	
	hosted e-shop: *Anything else?*
The user: *No thanks!*	
	hosted e-shop: *Your order is registered: a quattro stagioni pizza which costs 6£ and is expected to be delivered in about 15 minutes.*
The user terminates the communication.	

The voice portal hosting system 2 may comprise a standard digital computer 2 connected both to a first telecommunication network 10, e.g. a PSTN, ISDN, GSM or TCP/IP-Network, and to a second telecommunication network 3, e.g. TCP/IP, PSTN, ISDN or X25 network, and comprising an internal memory (not shown) in which a computer program product 26 can be directly loaded for performing the method steps of the invention when said product is run on said computer.